THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Clockwork Automata, Artificial Intelligence, and Why the Body of the Author Matters

Citation for published version:
Crosthwaite, P 2011, Clockwork Automata, Artificial Intelligence, and Why the Body of the Author Matters. in D Coleman & H Fraser (eds), *Minds, Bodies, Machines, 1770-1930.* Palgrave Macmillan, pp. 84-103. <http://www.palgrave.com/page/detail/minds-bodies-machines-17701930-deirdre-coleman/?K=9780230284678>

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Minds, Bodies, Machines, 1770-1930

Publisher Rights Statement:
© Crosthwaite, P. (2011). Clockwork Automata, Artificial Intelligence, and Why the Body of the Author Matters. In D. Coleman, & H. Fraser (Eds.), Minds, Bodies, Machines, 1770-1930. (pp. 84-103). Palgrave Macmillan. reproduced with permission of Palgrave Macmillan.

This extract is taken from the author's original manuscript and has not been edited. The definitive, published, version of record is available here: http://www.palgrave.com/page/detail/minds-bodies-machines-17701930-deirdre-coleman/?K=9780230284678.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 11. Nov. 2025

Crosthwaite, P 2011, 'Clockwork Automata, Artificial Intelligence, and Why the Body of the Author Matters'. in D Coleman & H Fraser (eds), *Minds, Bodies, Machines, 1770-1930*. Palgrave Macmillan, pp. 84-103.

4. Clockwork Automata, Artificial Intelligence, and Why the Body of the Author Matters

Paul Crosthwaite

On 9 June 1949, Geoffrey Jefferson, Professor of Neurosurgery at the University of Manchester, marked his receipt of the prestigious Lister Medal from the Royal College of Surgeons by addressing the College's members on the topic of 'The Mind of Mechanical Man'. Jefferson summarized his assessment of the prospects for mechanical consciousness with these words:

> Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain – that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.[i]

This passage articulates a set of principles that would come to guide much of the research in the fields that we now know as artificial intelligence (AI) and artificial life (AL): that there is a fundamental difference between the mere rote processing of data and the self-conscious awareness and understanding of what is being processed; that authentic consciousness is coloured by shifting emotional, affective, and libidinal states; and that this rich psychic reality finds its privileged expression in acts of artistic creation. This essay explores how these conceptions have been channelled into attempts to design computer programmes capable of producing original works of literature. It does so, however, by drawing parallels between recent research by computer scientists into the possibility of constructing artificial authors and the elaborate clockwork writing automata produced by European craftsmen in the late eighteenth to mid-nineteenth centuries. Building on the work of Jessica Riskin and others, I suggest that for all their comparative lack of sophistication, the clockwork

writers of the eighteenth and nineteenth centuries anticipate – by virtue of their embodiment in articulated, humanoid form – an emerging paradigm in the field of artificial intelligence that presents tantalizing possibilities for the development of machine creativity. This new research agenda has, in turn, surprising and profound consequences for literary criticism and theory today.

Writing Automata and Romantic Authorship

Between 1768 and 1844, the Swiss watchmakers Pierre and Henri-Louis Jaquet-Droz and Jean-Frédéric Leschot, the Jaquet-Droz's protégé Henri Maillardet, and the French illusionist Jean-Eugène Robert-Houdin constructed a series of startlingly lifelike writer-figures (figures 1-4).

FIGURE 1. Left to right: 'The Draftsman', 'The Musician', and 'The Writer' (constructed 1768-1774) by Pierre and Henri-Louis Jaquet-Droz and Jean-Frédéric Leschot. Photograph by Rama, Wikimedia Commons, Cc-by-sa-2.0-fr.

FIGURE 2. ‘The Writer’ by Pierre and Henri-Louis Jaquet-Droz and Jean-Frédéric Leschot. Photograph by Rama, Wikimedia Commons, Cc-by-sa-2.0-fr.

FIGURE 3. Writer-Draughtsman (1805) by Henri Maillardet. From the Historical and Interpretive Collections of The Franklin Institute, Inc., Philadelphia, PA.

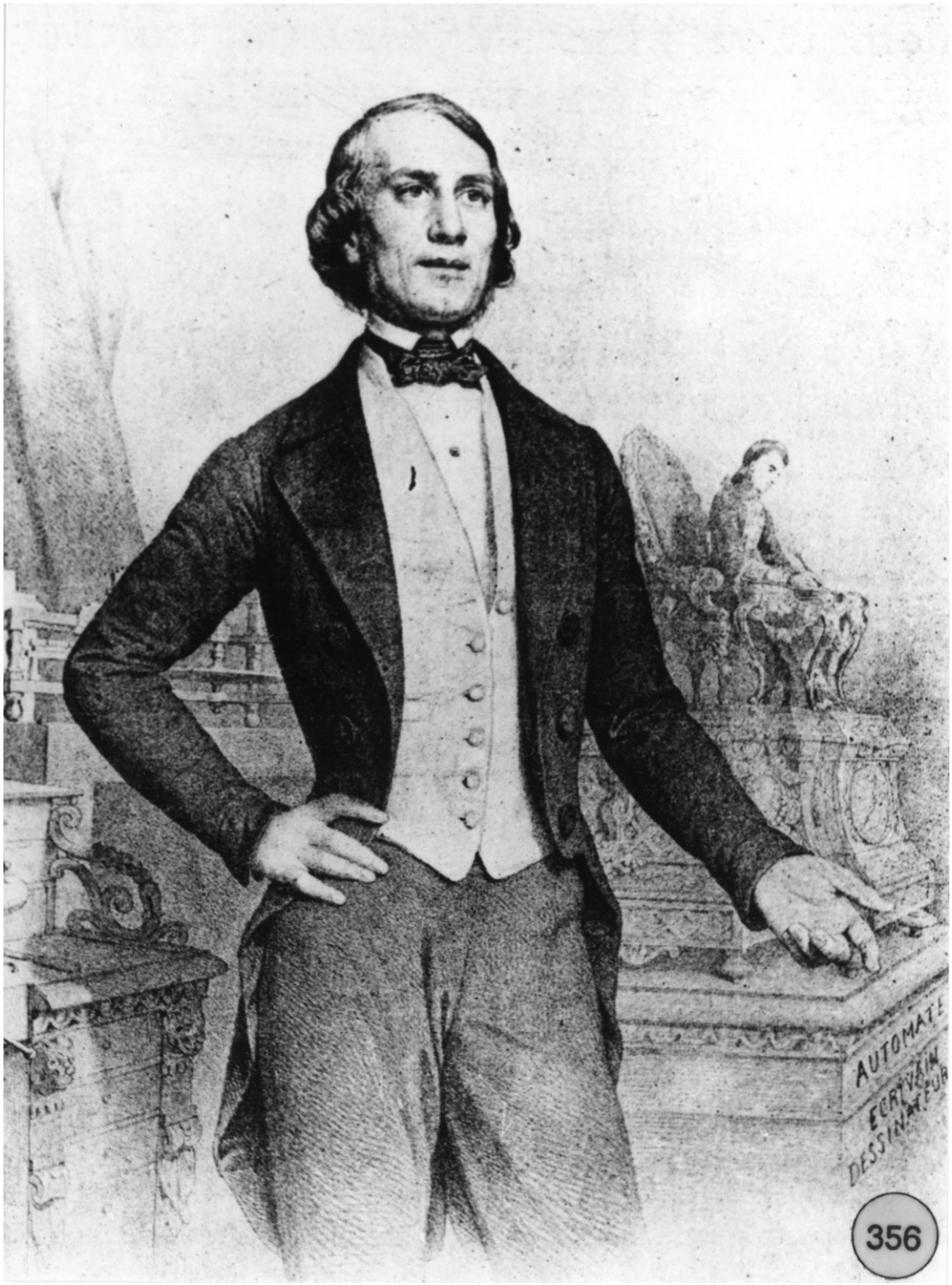

FIGURE 4. Jean-Eugène Robert-Houdin with his Writer-Draughtsman (1844; destroyed by fire 1865?) in background. Arts Collection, State Library of Victoria, Australia.

Seated at their desks and equipped with quills, they astounded viewers by inscribing documents set before them with signatures or short poems. The texts to be written by the Jaquet-Droz and Maillardet devices were ‘coded’ by selecting letters on a wheel and/or by setting a series of cams. Turned by clockwork motors, these components

drove complex systems of levers and rods, which guided the movements of the figures' hands over the page. Robert-Houdin's writer is thought to have been destroyed by a fire in 1865, and so has not been subjected to expert examination, but it seems likely that, in common with other automata built by the great illusionist, it would have combined elements of the sophisticated mechanisms utilized by the Jaquet-Drozes and Maillardet with hidden levers or pedals controlled by a human operator.[ii] Jessica Riskin locates this difference in the construction of these automata in the context of a shift from an eighteenth-century ethos of simulation (which sought to replicate, as accurately as possible, the mechanics of physiological processes themselves) to a nineteenth-century culture of analogy (which was content with devices that merely presented an outward semblance of such underlying processes).[iii] As Riskin acknowledges, however, even the device she identifies as the principal embodiment of the philosophy of simulation – Jacques Vaucanson's digesting duck (first exhibited 1738) – employed a crude fraud, rather than the elaborate technological architecture proclaimed by its maker, to achieve its effects (a fraud which was eventually revealed by none other than Robert-Houdin).[iv] Without discarding Riskin's distinction between simulation and analogy, then, it is nonetheless evident that throughout the eighteenth and early nineteenth centuries, automaton-makers were united in manipulating the outwardly visible form and function of their machines in order to effect an illusion of interior organic process, irrespective, to some degree at least, of the actual process employed. In the case of the writing automata of the period, the intended illusion, of course, was the uncanny impression that the figures were not simply rehearsing a series of predetermined actions, but independently generating their own writings. Self-evidently, moreover, this illusion was reliant on the convincingly humanoid appearance and behaviour of the writers, while it would have been dispelled by the exposure of the mechanism (of whatever kind) propelling their motions. As a character in E.T.A Hoffmann's 'Die Automaten' (1821) says of an automaton he has observed: 'The outward form … [of the figure] has been cleverly selected. Its shape, appearance, and movements are well adapted to occupy our attention in such a manner that its secrets are preserved and to give us a favourable opinion of the intelligence which gives the answers'.[v] The impression of autonomous agency thus conveyed could be quite profound. John Tresch has recently charted the varied and ambiguous responses that such devices elicited when they were displayed to the public in theatres and exhibition halls. While some, generally more

educated, viewers approached them as amusing novelties or impressive displays of technical ingenuity, others were willing to entertain the notion that they possessed some genuine flash of vitality.[vi] It is striking, though, that in accounts of public displays of these figures, even the most sober witnesses testify to at least a fleeting illusion of spontaneous creativity. As I have suggested, this effect was dependent on the machines' status as androids (humanoid automata) with anatomically proportionate physiques and naturalistic physiognomies; and this was for two reasons, both of which bear on prevailing models of subjectivity and authorship in the period.

Much discussion surrounds the cultural, philosophical, and aesthetic meanings and associations of these and other, contemporaneous, automata. Simon Schaffer, for example, identifies them as literal embodiments of a tradition of seventeenth- and eighteenth-century European thought that viewed humans as purely mechanical beings, a tradition whose origins lie in René Descartes' philosophy of animal existence, and which, in extending Descartes' claims to humans, reached its culmination in the French Enlightenment materialist ideas of Denis Diderot and, pre-eminently, Julian Offray de La Mettrie. Its definitive statement is La Mettrie's 1747 treatise *Machine Man*, with its conclusion that 'man is a machine and … there is in the whole universe only one diversely modified substance'.[vii] As Riskin emphasizes, though, even the most mechanistic of French Enlightenment philosophies found no contradiction in celebrating humanity's capacity for 'sentiment' and 'sensibility' – for feeling, emotion, passion, and expression;[viii] the attempts of automaton-makers to employ mechanical means in order to build figures that exhibited all the animation and vitality of human beings was, she suggests, wholly consistent with this outlook. The importance of a language of sentiment and sensibility throughout the eighteenth and early nineteenth centuries is, furthermore, just one of many continuities that have led critics and cultural historians to question conventional divisions between Enlightenment and Romantic world-views.[ix] Over a period exactly contemporaneous with the production of the clockwork writers discussed above, this tradition effloresced into the Romantic discourse of authorship, with its vision of literature as 'fundamentally expressive of a unique individuality' and defined by 'originality' and the 'conscious intention of the autonomous subject'.[x] Just as Romanticism privileged the expressive capacity of the writer, so, as Christopher Keep argues, the ability of an automaton to write suggests, more strongly than any other function it might possess,

> the presence not of a program but of a person, one whose actions are the free and spontaneous expressions of some deep reserve of selfhood, an inwardness or depth of being which is capable of reflecting on itself as self. The very appearance of writing … is always marked by the trace or outline of a living presence, the unique individual who is both the source and origin of the enunciative act.[xi]

Vivian Sobchack also interprets these machines, like similar devices now marketed as children's toys, as dramatizing a Romantic conception of authorship through the act of writing – specifically through the act of writing *by hand*. Handwriting, she remarks, 'is always … *auratic* insofar as it is enabled not just by a material body but by a *lived body* that, however regulated, cannot avoid inscribing its singular intentionality in acts and marks of *expressive improvisation*'.[xii]

The author of Romantic theory is a notoriously contradictory being, however, and if the convincingly anthropomorphous performances of these mechanical writer-figures partake of an expressivist or idealist vision of literary creativity, in which the imagination is granted an autopoietic status, they equally resonate with an empiricist model that stresses the writer's constitutive receptivity to the dynamics of the external world. As Riskin notes, the automaton-makers of the eighteenth and nineteenth centuries did not attempt to replicate the action of the five senses;[xiii] again, though, the verisimilar appearance of their machines succeeded in conveying to viewers an impression of responsiveness and alertness. In the case of the Jaquet-Drozes, even the choice of the figure's footwear (or lack of it) was designed to suggest an acute sensitivity to the environment (see figure 1). As Gaby Wood puts it,

> some inventors intended their objects to be artificial forms of an eighteenth-century ideal – the child as a blank slate, the purest being. The Jaquet-Droz figures conduct their marvellous activities barefoot, illustrating a belief, held by their contemporary Jean-Jacques Rousseau, that children would learn more freely if unhampered by shoes.[xiv]

As has often been noted, the exemplary synthesis of the expressivist-idealist and empiricist currents in Romantic aesthetics is found in William Wordsworth's 'Tintern Abbey' (1798).[xv] Wordsworth had no enthusiasm for automata,[xvi] but his exultant

celebration of 'all the mighty world / Of eye and ear, both what they half-create, / And what perceive' (106-08) precisely delineates the spectrum of faculties that the clockwork writers of his day were designed to give the impression of possessing. I will return to Wordsworth's poem later, in light of the connections I now wish to draw between these writer-figures and some recent developments in artificial intelligence research. If, in the late eighteenth and early nineteenth centuries, the writing automaton's humanoid and (partially) mobile construction was technically incidental, and served merely to generate an illusion of sentience and perception, in our own era such material embodiment is increasingly viewed as essential to the creation of genuine artificial intelligence.

Meaning and Embodiment in Machine-Generated Literature

In November 1928, the Franklin Museum in Philadelphia took delivery of the damaged and disassembled components of a brass clockwork machine. The donors, in whose family the device had resided for several generations, understood it to have once been capable of writing messages and drawing pictures, and had some notion of it being the work of the German inventor Johann Nepomuk Maelzel. After an engineer at the Institute had painstakingly repaired the device, it was equipped with a fountain pen and set in motion. It promptly inscribed four drawings and three poems, signing the last with the flourish, '*Ecrit par L'Automate de Maillardet*'.[xvii] This wonderfully eerie story – a signal manifestation of that effect we have come to call the uncanny – perfectly allegorizes the historical dynamic I wish to explore, in which the fidelity of later generations to the embodied form of early clockwork automata permits those figures to address us, with arresting directness, across the centuries.

Riskin notes that the conviction in this earlier moment 'that life, consciousness, and thought were essentially embodied in animal and human machinery has striking parallels in current Artificial Intelligence'.[xviii] As she observes, the notion that intelligence must be 'physically grounded' is the central principle of the sub-discipline of AI known as artificial life (AL). She cites the pioneering work of Rodney Brooks, director of the Artificial Intelligence Lab at MIT, who 'has left behind the purely software model of AI, and instead builds robots with sensors and feedback loops, giving them vision, hearing, and touch'.[xix] The idea that, as Susan Blackmore puts it, 'mind can be created only by interacting in real time with a real

environment' is of particular significance in the branches of AI that attempt to equip machines with a grasp of language.[xx] Perhaps the most significant challenge to conventional, box-bound AI programmes in this regard is the argument – made, most influentially, by the philosopher of mind John Searle – that such systems will never possess genuine linguistic ability because they are condemned to an existence that lacks 'intentionality'; that is, while they may be able to follow (and even, with the advent in the 1980s of artificial neural networks, progressively learn or internalize) syntactic rules, they can have no understanding of what the symbols they manipulate according to these rules are 'about', what they signify or refer to, what they *mean*. The 'bottom-up' approach advocated by proponents of 'embodied cognition' seeks to redress the problem of intentionality by more closely replicating the processes of human language acquisition:

> As human infants develop linguistic competence, they learn not only how to describe objects but also how to describe and express intentional relations such as wants, likes, and dislikes, intentional relations that were experienced by the infant before they could be cast in linguistic form. In this way language and cognition elaborate on previously experienced nonlinguistic and noncognitive (i.e., bodily) intentional relationships. An infant's experiences of noncognitive intentional relationships provide the foundation necessary for the cognitive life it will later enjoy. Human beings do not suffer from the symbol grounding problem … precisely because we are embodied….
>
> By building robots that interact with the environment prior to linguistic competence … an embodied approach to AI provides the necessary foundation for higher cognition.[xxi]

This new paradigm has not so far been extended into the domain of machine-generated literature. The potential of embodied cognition for this field can be most readily grasped, however, by considering the software-based systems that currently exist. I will focus on the two most prominent examples: the Cybernetic Poet designed by the American inventor Ray Kurzweil, and BRUTUS, a short story-writing programme developed by Selmer Bringsjord, David Ferrucci, and a team of computer scientists at Rensselaer Polytechnic Institute in New York State. Given their significance for questions of authorship and textual meaning, and the considerable

scholarly interest in other forms of electronic literature such as hypertext and interactive fiction, these projects have received remarkably little attention in literary studies.[xxii]

Kurzweil's Cybernetic Poet works by 'reading' – as its inventor puts it – poems by an author or authors and constructing a model of their work. The programme develops algorithms that allow it to imitate the style, metre, poetic structure, and vocabulary characteristic of the author(s). Kurzweil states that the resulting 'poems are in a similar style to the author(s) originally analyzed but are completely original new poetry'.[xxiii] The following are examples of poems generated by Kurzweil's programme, which can be accessed on his web site:

Soul
A haiku written after reading poems by John Keats and Wendy Dennis

You broke my soul
the juice of eternity,
the spirit of my lips.

And Pink In Sex
A haiku written after reading poems by Walt Whitman

Ages and pink in Sex,
Offspring of the
voices of all my Body.

Bringsjord and Ferrucci's BRUTUS produces short stories of up to 500 words using a sophisticated 'story grammar' architecture, which can handle character, setting, plot development, and the other basic elements of prose narrative. The first incarnation of BRUTUS took eight years to develop. Its designers concentrated on equipping the programme with the ability to write stories centred around the theme of betrayal, since this was one aspect of human experience that, they reasoned, could be logically tabulated, in contrast to more diffuse emotional phenomena such as love, fear, or regret. One of BRUTUS' stories, entitled simply 'Betrayal', begins like this:

> Dave Striver loved the university. He loved its ivy-covered clocktowers, its ancient and sturdy brick, and its sun-splashed verdant greens and eager youth. He also loved the fact that the university is free of the stark unforgiving trials of the business world – only this *isn't* a fact: academia has its own tests, and some are as merciless as any in the marketplace. A prime example is the dissertation defense: to earn the PhD, to become a doctor, one must pass an oral examination on one's dissertation. This was a test Professor Edward Hart enjoyed giving.[xxiv]

Literature generators such as the Cybernetic Poet and BRUTUS invite comparison with an avant-garde tradition of 'machine writing' that extends back to the beginning of the twentieth century. As Brian McHale demonstrates in an important recent survey, this tradition encompasses such varied innovations as the procedural compositional techniques employed by Raymond Roussel, the literary 'games' developed by the Surrealists, the cut-up strategies pioneered by the Dadaists and later adopted by William Burroughs, the author-computer collaborations undertaken by Charles O. Hartman, and the aleatory and/or arbitrarily rule-bound methods pursued by Louis Zukofsky, Jackson Mac Low, the Language poets, and the OuLiPo circle. While only a few of these figures and movements utilize actual machines, their texts are all examples of machine writing in the sense of that they are 'not "freely" composed but produced by the operation of mechanical techniques for generating and/or manipulating bits of language'.[xxv] McHale argues that these texts can be best compared to one another 'in terms of the relative proportions of writer to machine participation in the composition of the text'.[xxvi] All literature, he argues, possesses a mechanical element to the extent that it imposes constraints of form, genre, length – and, specifically in the case of poetry, rhyme and metre – on the expressive capacity of the author.[xxvii] The spectrum of machine writing as such, however, ranges from texts in which a set of more-or-less mechanically produced materials are submitted to heavy postprocessing on the part of the writer (as in the novels of Raymond Roussel), through to the practice of a figure such as Jackson Mac Low, which entirely delegates the outcome of a particular procedure or programme, and even in certain instances the choice of the procedure itself, to mechanical permutation.[xxviii] Kurzweil's and Bringsjord and Ferrucci's generators would find themselves at the latter pole of McHale's typological scale: they produce their texts

within certain pre-programmed parameters, but the process is not otherwise subject to human intervention or interference. While avant-gardist utilizations of mechanical techniques consistently present themselves as strategic subversions of the autonomous, expressive author who looms so large in Romantic and post-Romantic aesthetic ideology, however, the Cybernetic Poet and BRUTUS projects seemingly aspire to endow the machine itself with those very qualities of sovereignty and creative genius. Tellingly, Kurzweil has a clear preference for poets in the Romantic tradition, including Keats and Whitman, as well as Blake, Byron, and Shelley (though the relation of his own mechanical poet to these titans is, of course, purely imitative). Bringsjord and Ferrucci, meanwhile, identify the benchmark of creativity with such giants of the canon as Dickens, Tolstoy, Joyce, Updike, and Morrison (whilst acknowledging that 'if $BRUTUS_n$, some refined descendant of $BRUTUS_1$, is to soon find employment at the expense of a human writer, in all likelihood it will be as an author of formulaic romance and mystery'[xxix]).

Regardless of the shortcomings of these programmes with respect to their designers' highest ambitions, and whatever we might think of the aesthetic qualities of the writings they produce, it is undeniable that their texts bear at least a passable resemblance to the literary forms they are designed to emulate. As Kathleen L. Komar says of a piece by Kurzweil's poet, 'if we did not know this [the text's mechanical provenance], we would undoubtedly count the poem as literature'.[xxx] What is equally clear, however (and what I take to be the grounds for Komar's equivocation), is that these programmes remain entirely bound by the problem of intentionality: they may be able to follow rules in such a way as to produce texts that meet the objective criteria for recognition as works of poetry or short fiction, but they have no apprehension of what these texts mean, or even that they *could* yield such a thing as meaning. The successes and failures of these projects are cast into sharp relief by two celebrated intellectual experiments, which are often hailed as inaugurating, respectively, the disciplines of artificial intelligence and modern literary theory: the eponymous test invented by the computer scientist Alan Turing to determine the existence of machine intelligence, and the experiment in literary response undertaken by the critic I.A. Richards under the banner of 'practical criticism'.

The 'Turing Test', first described in the seminal paper 'Computing Machinery and Intelligence' (1950), consists of a scenario in which a human interrogator poses questions to two concealed interlocutors in an attempt to determine which is a human

and which a machine; if the machine can persuade the interrogator that it is the human party, Turing reasoned, then it can be legitimately deemed intelligent. Both the Cybernetic Poet and BRUTUS have been submitted to variations on the Turing Test, in which readers attempted to distinguish the programmes' texts from those by human writers.[xxxi] The thirteen adults and three children to whom Kurzweil administered his test correctly attributed the poems they read at a rate of 63 per cent and 48 per cent, respectively.[xxxii] Meanwhile, 25 per cent of the two thousand web visitors who read a piece by BRUTUS alongside four stories by human writers successfully identified the machine-authored text.[xxxiii] The fact that in these (albeit only semi-scientific) tests readers succeeded in distinguishing between human- and computer-generated writings at a rate not significantly better than chance lends empirical weight to the assertion that the two programmes are capable of imitating the conventions of their assigned genres with a considerable degree of credibility. Paradoxically, however, when the products of these programmes are considered in light of an experiment that insists on curbing considerations of authorial identity in favour of concentrated interpretation of the texts themselves, the irrevocable alterity of their computational origins – seemingly elided in these quasi-Turing Tests – reasserts itself.

I.A. Richards' *Practical Criticism* (1929) describes an initiative undertaken at Cambridge University in the 1920s, whereby groups of readers, predominantly undergraduates studying English, were issued with poems by a range of authors – contemporary, canonical, and minor – which had been stripped of personally and historically identifying details; the readers were invited to submit written responses or 'protocols' in which they recorded their reflections on these materials. Richards' realisation, on the basis of these often misconceived submissions, that literary criticism needed to develop a considerably more rigorous and systematic methodology prepared the ground for many of the theoretical innovations of the succeeding decades. His pedagogical exclusion of biographical and historical data so as to focus the reader's attention on the words on the page would prove to be particularly significant for the Anglo-American New Critics of the 1940s and '50s. Its influence is notably discernible in William K. Wimsatt and Monroe C. Beardsley's famous refutation of the so-called intentional fallacy on the grounds that 'the design or intention of the author is neither available nor desirable as a standard for judging the success of a work of literary art'.[xxxiv] Despite the very different intellectual coordinates of the two traditions, affinities can also be detected between the vision of

the text as an autonomous artefact pioneered by Richards and formalized by the New Critics and the radical anti-authorialism and anti-intentionalism of certain strands of French poststructuralist theory, positions most vividly articulated in Roland Barthes' notorious 1967 essay 'The Death of the Author'. There are, for example, clear resonances with Wimsatt and Beardsley in Barthes' claim that in traditional criticism 'the *explanation* of a work is always sought in the man or woman who produced it, as if it were always in the end … the voice of a single person, the *author* "confiding" in us'.[xxxv] Similarly, like the New Critics, Barthes is not only resistant to critical methods that seek to ground meaning in the figure of the author, but also to those which appeal to the author's 'hypostases': namely, 'society, history, psyche, liberty', entities which are imagined as dwelling 'beneath the work' and which, once 'found', explain it.[xxxvi]

The most significant challenge to the New Critical and poststructuralist assaults on authorial intention came in 1982 with Steven Knapp and Walter Benn Michaels' polemical essay 'Against Theory'. Knapp and Michaels argue that meaning is inextricable from intention: however much material marks may resemble familiar textual signifiers, they cannot be understood as meaningful unless they are intentionally inscribed. They invite the reader to imagine encountering what appear to be lines of poetry etched on the beach: if we count these marks 'as nonintentional effects of mechanical processes (erosion, percolation, etc.)' then to treat them as meaningful would be an invalid projection of agency onto merely contingent phenomena.[xxxvii] On the basis of this argument, Knapp and Michaels make the bold claim that since there can be no meaning without intention, 'the meaning of a text is simply identical to the author's intended meaning'.[xxxviii] Given this, 'theory' – by which they mean 'the attempt to govern interpretations of particular texts by appealing to an account of interpretation in general' – is a misguided enterprise that should be abandoned. Though few scholars have been willing to accept Knapp and Michaels' arguments wholesale, it is nonetheless apparent in retrospect that, as Reed Way Dasenbrock remarks, '"Against Theory" and the controversy it generated helped usher in the "post-theoretical" era we now seem to be in'[xxxix] – 'post-theoretical' to the extent, at least, that debates over the appropriate hermeneutic or interpretive protocols for textual analysis no longer have the centrality in critical practice that they once had. In their wake, the prevailing tendency has been away from a text-centred focus on the disentanglement of meaning and towards a re-embedding of those

meanings within the kind of extra-textual fields that the New Criticism and the 'high theory' of poststructuralism both, in their different ways, sought to bracket out. A return to history, materiality, referentiality, the experiential, the bodily, and the real is evident across an array of recently emergent or re-invigorated critical movements, ranging from new historicism and cultural materialism to Marxism, postcolonialism, feminism, queer theory, trauma studies, and ecocriticism. Michaels, in particular, has made significant contributions to this contextual or historicist turn in literary studies (most notably his major new historicist work *The Gold Standard and the Logic of Naturalism* [1987]), but if the 'After Theory' controversy played a part in paving the way for this shift, it did so more through its challenge to the dominance of a critical paradigm whose interests tended to exclude extra-textual concerns than through any positive endorsement of those concerns themselves; indeed, Knapp and Michaels insist that they make no claims at all 'about what should count as evidence for determining the content of any particular intention' (intention for them, of course, being synonymous with meaning).[xl] Unexpectedly, the field of embodied cognition – and, more distantly, the clockwork writing automata that so suggestively anticipate its interests – indicate ways in which the argument of 'After Theory' might be extended and modified so as to establish a compelling ontological legitimation for the expanded horizons of recent critical study.

The grounds for this legitimation begin to become clear when one considers how a successor of one of Richards' students might respond if asked to write a 'protocol' on a suitably anonymized text by the Cybernetic Poet or BRUTUS. The reader would no doubt be able to give some account of the basic, literal sense of the piece, and might also succeed in tracing some credible patterns of imagery or paths of thematic development, but, once informed of its origins, they would be likely to feel that the exercise had been in some way profoundly futile. As P.D. Juhl observes, there is 'something odd about *interpreting*' a 'computer poem'.[xli] Accordingly, McHale describes the 'resentment' that 'anyone who has introduced [interactive, machine-mediated, or machine-generated] poetry to students knows'.[xlii] Insofar as the function of such avant-gardist strategies is precisely to challenge the reduction of literary reception to a pure matter of determining meaning, McHale's implicit impatience with his students is understandable enough, but, equally, if the urge to decipher familiar, apparently intelligible signs is not simply a convention of certain forms of literary training but an integral element of our very species-being, then the students'

resentment is equally excusable. Indeed, to return to the hypothetical example of an exercise in practical criticism being performed on a machine-generated text, the reader's response – which is in this case wholly predicated on the establishment of meaning – would not only feel futile, but *would* be futile, since it would consist of a mere encounter with the 'nonintentional effects of mechanical processes',[xliii] from which it is as perverse to read off meaning as it is instinctive to do so.

The Cybernetic Poet, BRUTUS, and other highly 'delegated' systems of machine writing truly are hypostases of Barthes' dead author: thoughtless, affectless, intentionless beings whose arbitrary manipulations of 'tissue[s] of quotations drawn from the innumerable centres of culture' function, as if by magic, to drain these textual fragments of their significatory power.[xliv] All that remains for their readers is the possibility of a delirious *dérive* across the smooth surface of the text, in pursuit not of interpretation or decipherment,[xlv] but of the sheer overwhelming *jouissance* evoked by the material signifier in its all geometric splendour. Despite the best efforts of Barthes and others – including, most notably, Susan Sontag and Gilles Deleuze and Félix Guattari[xlvi] – the notion of a genuinely non-interpretive aesthetics remains less a critical programme, however, than an intriguing thought experiment, one which has in fact only served to demonstrate the inherently interpretive character of every critical statement.

Meaning, then, is the uncircumventable object of reading, and meaning, as we have seen, can be guaranteed only by the agency of an intentional being. Of course, the Cybernetic Poet's 'Soul', BRUTUS' 'Betrayal', or Jackson Mac Low's 'Call Me Ishmael' were not, as in Knapp and Michaels' example, engraved on the beach by some cosmically improbable accident. They each originated, instead, in the actions of a programmer or designer, who presumably had some understanding of the rules and symbols he or she selected for mechanical processing, and some aspiration that, once initiated, the programme would combine these materials in such a way as to produce textual outputs intelligible to a human reader. As Kathleen L. Komar remarks with regard to the Cybernetic Poet, 'the initial reading experience' of Kurzweil and his programming colleagues 'informs the programs they write to create new texts that will produce a similar experience for the reader'.[xlvii] A marginal degree of intentionality, and thus of meaning, can be recuperated in these instances, therefore, but only by appealing to the human agent or agents without whose initiating role no such texts would exist. Why is it, then, that writings by John Donne, Edna St. Vincent

Millay, or the Reverend G.A. Studdert Kennedy, which I.A. Richards invited his Cambridge classes to respond to in the 1920s, enjoy an intentional and semantic plenitude inevitably withheld from the algorithmically-generated text, or available only to the extent that it is guaranteed by the activity of a human programmer? The answer – as the clockwork automata of the eighteenth and nineteenth centuries long ago hinted, and as artificial intelligence has recently demonstrated – is that intentionality can arise only from embodied existence in the referential realm of material objects and relations.[xlviii] Wordsworth thus thematizes the very ontological conditions of possibility of his own poetry when, in 'Tintern Abbey', he strives to summon a vision of his younger self driven by presymbolic, animalistic impulsions to range across the as yet undifferentiated world of phenomena:

> changed, no doubt, from what I was, when first
> I came among these hills; when like a roe
> I bounded o'er the mountains, by the sides
> Of the deep rivers, and the lonely streams,
> Wherever nature led;
> …….
> For nature then
> (The coarser pleasures of my boyish days,
> And their glad animal movements all gone by,)
> To me was all in all. – I cannot paint
> What then I was. The sounding cataract
> Haunted me like a passion: the tall rock,
> The mountain, and the deep and gloomy wood,
> Their colours and their forms, were then to me
> An appetite: a feeling and a love,
> That had no need of a remoter charm,
> By thought supplied, or any interest
> Unborrowed from the eye. – That time is past. (67-71; 73-84)

To quote Jacques Lacan, it is only from this organic, infantile union with the 'the entirety of things, ... the totality of the real' that the speaker's capacity for linguistic

reflection on his condition can emerge, inscribing 'on the plane of the real this other plane, which we here call the plane of the symbolic'.[xlix]

Those modes of literary analysis that attempt to separate literary texts out from the spatio-temporal manifold in which they are situated therefore paradoxically exclude the very phenomena that make meaning, and thus criticism itself, possible. This being so, the incorporation of these phenomena – whether they be, say, a soaring rock formation, a ruined religious building, the scars carved on the landscape by the rhythms of industrialization, or the violent upheavals on the streets of revolutionary Paris – into our reflections on literary meaning becomes less a matter of preference than of necessity. The state of embodied intentionality that subtends literary meaning demands, then, a wider consideration of the world through and in which this state develops; but the overdetermined nature of the subject's worldlihood rules out any endorsement of Knapp and Michaels' claim (in the face of the anti-intentionalism of the New Criticism and poststructuralism) that the 'the meaning of a text is simply identical to the author's intended meaning'.[l] If the embeddedness of the human subject in the material conditions of life on earth permits the emergence of its capacity for meaning-making, then, in a recursive movement, it is the privileged manifestation of this capacity – literature – that most powerfully crystallizes these more-or-less contingent and impersonal conditions into meaningful, symbolic form. Such is the virtually infinite variety of these conditions, however, that this intervention on the part of the writer constitutes the coming into being of a field of semantic potential in which meaning may be almost inexhaustibly sought and found. The literary act is the performative announcement of an intention to mean, not the inscription of a singular intended meaning.

Prolegomenon to a Robot Literary History

Two days after Geoffrey Jefferson delivered the Lister Oration quoted at the beginning of this essay, in which he cast doubt on the likelihood of a machine ever genuinely replicating the human composition of a sonnet, his colleague at the University of Manchester, Alan Turing, was quoted in *The Times* as saying,

> I do not see why it [a computer at the University] should not enter any one of the fields normally covered by the human intellect, and eventually compete on equal terms. I do not think you can even draw the line about sonnets though the

> comparison is perhaps a little bit unfair because a sonnet written by a machine will be better appreciated by another machine.[li]

By the late 1980s, rudimentary computer-generated poetry was well established. In an essay on a notable early programme, RACTER, Christian Bök speculates that 'the poets of tomorrow are likely to resemble programmers, exalted, not because they can write great poems, but because they can build a small drone out of words to write great poems for us'. He continues: 'What have we to lose by writing poetry for a robotic culture that must inevitably succeed our own?... We may have to consider this heretofore unimagined, but nevertheless prohibited, option: writing poetry for inhuman readers, who do not yet exist, because such aliens, clones, or robots have not yet evolved to read it'.[lii] Casting an eye towards this far future, Michael L. Johnson wonders, 'what forms beyond the human could evolve, what new kinds of difficult beauty?... The rise of silicon "life", silicon intelligence: a Promethean act of technology and language. One may imagine silicon entities floating through deep space, manipulating signifiers beyond human ken'.[liii]

The perspective posited in these quotations is articulated at length by Manuel De Landa in his extraordinary book *War in the Age of Intelligent Machines* (1991), which invites us to imagine a future class of

> specialized 'robot historians' committed to tracing the various technological lineages that gave rise to their species. And we could further imagine that such a robot historian would write a different kind of history than would its human counterpart.... The robot historian ... would hardly be bothered by the fact that it was a human who put the first motor together: for the role of humans would be seen as little more than that of industrious insects pollinating an independent species of machine-flowers that simply did not possess its own reproductive organs during a segment of its evolution.[liv]

As the remarks by Turing, Bök, and Johnson suggest, a robot literary history would likely see human beings and their aesthetic interests as similarly marginal to its narrative. What seems increasingly clear, however, is that information processing machines will only conceivably develop the sentience necessary for a literary culture of their own by escaping the prison of nonintentionality; and their only possibility of

achieving this is by emulating humans to the extent, at least, of ceasing to dwell in grey boxes on laboratory desks, and emerging, instead, as embodied creatures free to explore the world they inhabit. Any such literary history would no doubt reserve privileged chapters for the clockwork writers of the late eighteenth and early nineteenth centuries, whose embodied forms so strikingly anticipate those of their robotic descendents, as well as the literature generators of our own present, which demonstrate the limits of an existing paradigm, and the necessity of new departures. Both moments also (although, of course, from our robot historian's perspective merely incidentally) cast new light on some central questions in what we will have to learn to call human literary history.

I am grateful to Melanie Waters and members of the audience at the 'Minds, Bodies, Machines' conference hosted by Birkbeck College, University of London in July 2007 for their valuable comments on earlier versions of this essay. I also wish to thank the Franklin Institute in Philadelphia and the State Library of Victoria for permission to reproduce images of two of the automata discussed above.

[i] Geoffrey Jefferson, 'The Mind of Mechanical Man', *British Medical Journal* (25 June 1949), 1110.

[ii] Simon During, *Modern Enchantments: The Cultural Power of Secular Magic* (Cambridge, MA: Harvard University Press, 2002), 121; Jessica Riskin, 'Eighteenth-Century Wetware', *Representations* 83 (2003), 118.

[iii] Riskin, 'Eighteenth-Century Wetware', 117-8.

[iv] Riskin, 'Eighteenth-Century Wetware', 104, 117.

[v] E.T.A. Hoffmann, 'Automata', in E.F. Bleiler (ed.), Major Alexander Ewing (trans.), *The Best Tales of Hoffmann* (Mineola, NY: Dover, 1967), 91.

[vi] On audience responses to automata in the period, see also During, *Modern Enchantments*, 121-3.

[vii] Simon Schaffer, 'Enlightened Automata', in William Clark, Jan Golinski and Simon Schaffer (eds), *The Sciences in Enlightened Europe* (Chicago: University of Chicago Press, 1999), 39.

[viii] Jessica Riskin, 'The Defecating Duck, or, the Ambiguous Origins of Artificial Life', *Critical Inquiry* 29.4 (2003), 611; and *Science in the Age of Sensibility: The Sentimental Empiricists of the French Enlightenment* (Chicago: University of Chicago Press, 2002), 1-4.

[ix] For an overview of such continuities, see chapter one of Aidan Day, *Romanticism* (London: Routledge, 1996), 7-78.

[x] Andrew Bennett, *The Author* (London: Routledge, 2005), 54, 57.

[xi] Christopher Keep, 'Of Writing Machines and Scholar-Gypsies', *English Studies in Canada* 29.1-2 (2003), 56. See also Bruce Mazlish who, alluding to a myth central to the Romantic movement, remarks that the Jaquet-Droz writer and similar automata extended the promise of 'creative, Promethean force' ('The Man-Machine and Artificial Intelligence', in Ronald Chrisley [ed.], *Artificial Intelligence: Critical Concepts* [London: Routledge, 2000], 138).

[xii] Vivian Sobchack, '"Susie Scribbles": On Technology, *Technë*, and Writing Incarnate', *Carnal Thoughts: Embodiment and Moving Image Culture* (Berkeley: University of California Press, 2004), 130, emphases in original.

[xiii] Riskin, 'Eighteenth-Century Wetware', 115-7.

[xiv] Gaby Wood, *Edison's Eve: A Magical History of the Quest for Mechanical Life* (New York: Knopf, 2002), xix-xx.

[xv] References to 'Tintern Abbey' are taken from Stephen Gill (ed.), *The Oxford Authors: William Wordsworth* (Oxford: Oxford University Press, 1984), 131-5.

[xvi] *The Prelude* (1805; published 1850) features a nightmare vision of a London fair, in which Wordsworth puns on the name of John Merlin, the proprietor of a mechanical museum in the West End (from Stephen Gill [ed.], *The Oxford Authors: William Wordsworth* [Oxford: Oxford University Press, 1984], 375-590). See Schaffer, 'Enlightened Automata', 137:

All moveables of wonder from all parts,
Are here,
……...
The Bust that speaks, and moves its goggle eyes,
The Wax-work, Clock-work, all the marvellous craft
Of modern Merlins, wild Beasts, Puppet-shows,
All out-o'-th'-way, far-fetched, perverted things,
All freaks of Nature, all Promethean thoughts
Of man; his dulness, madness, and their feats,
All jumbled up together to make up
This Parliament of Monsters. (VII.680-81; 685-92)

[xvii] Charles Penniman, 'Maillardet's Automaton', *The Franklin Institute*, http://www.fi.edu/learn/scitech/automaton/automaton.php?cts=instrumentation, 30 March, 2009.

[xviii] Riskin, 'Eighteenth-Century Wetware', 116.

[xix] Riskin, 'Eighteenth-Century Wetware', 116.

[xx] Susan Blackmore, *Consciousness: An Introduction* (London: Hodder and Stoughton, 2003), 189.

[xxi] Lewis A. Loren and Eric Dietrich, 'Merleau-Ponty, Embodied Cognition and the Problem of Intentionality', *Cybernetics and Systems: An International Journal* 28 (1997), 355-6.

[xxii] Notable exceptions are Kathleen L. Komar on Kurzweil's poet ('Candide in Cyberspace: Electronic Texts and the Future of Comparative Literature', *Comparative Literature* 59.3 [2007], xiii-xiv); Norbert Bachleitner, also on Kurzweil's poet, as well as several comparable projects ('The Virtual Muse: Forms and Theory of Digital Poetry', *Theory into Poetry: New Approaches to the Lyric*, Eva Müller-Zettelmann and Margarete Rubik [eds.] [Amsterdam: Rodopi, 2005], 316-28; and Christian Bök and Josef Ernst on an earlier, more rudimentary programme, William Chamberlain and Thomas Etter's RACTER ('The Piecemeal Bard is Deconstructed: Notes Toward a Potential Robopoetics', *Object* 10 [2002]; 'Computer Poetry: An Act of Disinterested Communication', *New Literary History* 23.2 [1992]). For a useful overview of current critical interests in the field of electronic literature, see Katherine N. Hayles, *Electronic Literature: New Horizons for the Literary* (Notre Dame: University of Notre Dame Press, 2008).

[xxiii] Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York: Viking, 1999), 163.

[xxiv] Selmer Bringsjord and David A. Ferrucci, *Artificial Intelligence and Literary Creativity: Inside the mind of BRUTUS, A Storytelling Machine* (Mahwah, NJ: Lawrence Erlbaum, 2000), 199-200, emphasis in original.

[xxv] Brian McHale, 'Poetry as Prosthesis', *Poetics Today* 21.1 (2000), 3.

[xxvi] McHale, 'Poetry as Prosthesis', 20.

[xxvii] McHale, 'Poetry as Prosthesis', 27-8.

[xxviii] McHale, 'Poetry as Prosthesis', 21-2.

[xxix] Bringsjord and Ferrucci, *Artificial Intelligence and Literary Creativity*, xii.

[xxx] Komar, 'Candide in Cyberspace', xiv.

[xxxi] Brian McHale proposes "a kind of Turing Test of poetry: If the machine imitates a human author's linguistic behaviour so perfectly that it can be mistaken for an author, then does it count as an author? And if not, why not?" ('Poetry as Prosthesis', 25 n. 22). I hope to offer a convincing answer to the latter question.

[xxxii] Ray Kurzweil, 'A Kind of Turing Test', http://www.kurzweilcyberart.com/poetry/rkcp_akindofturingtest.php3, 30 March, 2009.

[xxxiii] Carl Sommers, 'By the Way; Inspiration or Computation', *New York Times* (28 Novermber, 1999), http://www.nytimes.com/1999/11/28/nyregion/by-the-way-inspiration-or-computation.html?n=Top%2FReference%2FTimes%20Topics%2FSubjects%2FW%2FWriting%20and%20Writers, 2 April, 2009.

[xxxiv] William K. Wimsatt and Monroe C. Beardsley, 'The Intentional Fallacy', in William K. Wimsatt (ed.), *The Verbal Icon: Studies in the Meaning of Poetry* (Lexington: University of Kentucky Press, 1954), 3.

[xxxv] Roland Barthes, 'The Death of the Author', in Stephen Heath (ed. and trans.), *Image-Music-Text* (London: Fontana, 1977), 143, emphasis in original.

[xxxvi] Barthes, 'The Death of the Author', 147. The best study of this tradition of attacks on authorial intention remains Seán Burke's classic *The Death and Return of the Author: Criticism and Subjectivity in Barthes, Foucault, and Derrida* (Edinburgh: Edinburgh University Press, 1992).

[xxxvii] Steven Knapp and Walter Benn Michaels, 'Against Theory', *Critical Inquiry* 8.4 (1982), 728. The poem is Wordsworth's lyric 'A Slumber Did My Spirit Seal' from the 1800 edition of *Lyrical Ballads*. Knapp and Michaels admit that their 'wave poem' example may seem 'farfetched' but suggest that there '*are* cases where the question of intentional agency might be an important and difficult one', such as the debate over whether computers can 'speak' (728-9, emphasis in original).
[xxxviii] Knapp and Michaels, 'Against Theory', 724.
[xxxix] Reed Way Dasenbrock, *Truth and Consequences: Intentions, Conventions and the New Thematics* (University Park, P.A.: Pennsylvania State University Press, 2001), 156. On the controversy generated by 'Against Theory' see W.J.T. Mitchell, *Against Theory: Literary Studies and the New Pragmatism* (Chicago: University of Chicago Press, 1985).
[xl] Steven Knapp and Walter Benn Michaels, 'A Reply to Our Critics', *Critical Inquiry* 9.4 (1983), 796.
[xli] P.D. Juhl, *Interpretation: An Essay in the Philosophy of Literary Criticism* (Princeton: Princeton University Press, 1980), 84, 85, emphasis in original; qtd. in Knapp and Michaels, 'Against Theory', 731.
[xlii] McHale, 'Poetry as Prosthesis', 25.
[xliii] Knapp and Michaels, 'Against Theory', 728.
[xliv] Barthes, 'The Death of the Author', 146.
[xlv] Barthes, 'The Death of the Author', 147.
[xlvi] See Susan Sontag, 'Against Interpretation', *Against Interpretation and Other Essays* (1966; New York: Picador, 2001); Gilles Deleuze and Félix Guattari, *Anti-Oedipus: Capitalism and Schizophrenia*, Robert Hurley, Mark Seem, and Helen R. Lane (trans.) (1972; London: Athlone, 1984).
[xlvii] Komar, 'Candide in Cyberspace', xiv.
[xlviii] Selmer Bringsjord, one of the designers of BRUTUS, acknowledges that in order to move towards real creativity, the programme 'would need not only to think mechanically in the sense of swift calculation (the forte of supercomputers like [the chess-playing] Deep Blue), it would also need to think experientially in the sense of having subjective or phenomenal awareness. For example, a person can think experientially about a trip to Europe as a kid, remember what it was like to be in Paris on a sunny day with an older brother, smash a drive down a fairway, feel a lover's touch, ski on the edge, or need a good night's sleep' (*Singularity Summit at Stanford* [Stanford University, 2006], http://sss.stanford.edu/others/selmerbringsjord/, 7 April, 2009).
[xlix] Jacques Lacan, *The Seminar I: Freud's Papers on Technique, 1953-54* (New York: Norton, 1988), 262.
[l] Knapp and Michaels, 'Against Theory', 724.
[li] Elizabeth Wilson, 'Imaginable Computers: Affects and Intelligence in Alan Turing', Darren Tofts, Annemarie Jonson, and Alessio Cavallaro (eds), *Prefiguring Cyberculture: An Intellectual History* (Cambridge, MA: MIT Press, 2004), 43.
[lii] Bök, 'The Piecemeal Bard is Deconstructed', 17.
[liii] Michael L. Johnson, *Mind, Language, Machine: Artificial Intelligence in the Poststructuralist Age* (Basingstoke: Macmillan, 1988), 15.
[liv] Manuel De Landa, *War in the Age of Intelligent Machines* (New York: Zone, 1991), 2-3.